

The European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation

^{§,1,2}Mbaye Tine, ^{§,2}Heiner Kuhl, ^{§,3,4}Pierre-Alexandre Gagnaire, ⁵Bruno Louro, ³Erick Desmarais, ⁵Rute S. T. Martins, ^{2,6}Jochen Hecht, ²Florian Knaust, ³Khalid Belkhir, ²Sven Klages, ¹Roland Dieterich, ¹Kurt Stueber, ⁷Francesc Piferrer, ³Bruno Guinand, ^{3,4}Nicolas Bierne, ⁸Filip A. M. Volckaert, ⁹Luca Bargelloni, ⁵Deborah M. Power, ^{§,3,4}François Bonhomme, ^{§,5}Adelino V. M. Canario and ^{§,1,2}Richard Reinhardt

[§]Contributed equally as first authors

[§]Corresponding authors (francois.bonhomme@univ-montp2.fr, acanario@ualg.pt, reinhardt@mpipz.mpg.de)

¹Max Planck Genome-centre Cologne, Carl-von-Linné-Weg 10, D-50829 Köln, Germany

²Max Planck Institute for Molecular Genetics, Ihnestrasse 63, D-14195 Berlin, Germany

³Institut des Sciences de l'Evolution (UMR 5554), CNRS-UM2-IRD, Place Eugène Bataillon, F-34095 Montpellier Cedex 5, France

⁴Station Méditerranéenne de l'Environnement Littoral, Université Montpellier 2, 2 Rue des Chantiers, F-34200 Sète, France

⁵CCMAR-Centre of Marine Sciences, University of Algarve, Building 7, Campus de Gambelas, 8005-139 Faro, Portugal

⁶BCRT, Charité-Universitätsmedizin Berlin, Augustenburger Platz 1, D-13353 Berlin, Germany

⁷Institut de Ciències del Mar, Consejo Superior de Investigaciones Científicas (CSIC), Passeig Marítim, 37-49, 08003 Barcelona, Spain

⁸Laboratory of Biodiversity and Evolutionary Genomics, University of Leuven, Charles Deberiotstraat 32, B-3000 Leuven, Belgium

⁹Dipartimento di Biomedicina Comparata e Alimentazione, Università di Padova, I-35124, Padova, Italy

Abstract

The European sea bass (*Dicentrarchus labrax*) is a temperate zone euryhaline teleost of prime importance for aquaculture and fisheries. This species is subdivided into two naturally hybridizing lineages, one inhabiting the north-eastern Atlantic Ocean and the other the Mediterranean and Black seas. Here, we provide a high-quality chromosome-scale assembly of its genome that shows a high degree of synteny with the more highly derived teleosts. We find expansions of gene families specifically associated with ion and water regulation, highlighting adaptation to variation in salinity. We further generate a genome-wide variation map through RAD-sequencing of Atlantic and Mediterranean populations. We show that variation in local recombination rates strongly influences the genomic landscape of diversity within and differentiation between lineages. Comparing predictions of alternative demographic models to the joint allele-frequency spectrum indicates that genomic islands of differentiation between sea bass lineages were generated by varying rates of introgression across the genome following a period of geographical isolation.

Introduction

The European sea bass (*Dicentrarchus labrax*) is a teleost fish found in the north-eastern Atlantic Ocean and throughout the Mediterranean and Black seas. This is an economically important species, and natural stocks are subject to intensive exploitation by professional and sport fisheries, thus raising future conservation and management issues. It is ranked fourth in European aquaculture production and although undergoing domestication for the last three decades, selective breeding programs are still in their infancy.

The European sea bass inhabits coastal waters, where reproduction occurs, but it can also enter brackish waters in estuarine areas and coastal lagoons, and occasionally rivers. Thus, the European sea bass is a euryhaline fish tolerating a wide range of salinities (0-60 psu)¹. Although several genes involved in the physiological response to variation in salinity have been identified², the genetic basis of its broad halotolerance remains unclear. Because adaptation to varying environments may also involve gene duplications³, obtaining the complete genome sequence of the European sea bass may help identify the genetic basis for adaptation to euryhalinity.

Atlantic and Mediterranean sea bass populations represent two genetically distinct lineages that naturally hybridize in the Alboran Sea⁴. Describing genomic variation patterns that arise from the interaction between these two diverging lineages has important implications for understanding the evolutionary forces at play during the diversification of marine species. In addition, a detailed picture of genomic variation within and between lineages will provide useful information to orientate domestication and assist conservation management of sea bass populations⁵.

Here, we produce a high-quality draft sequence of the European sea bass genome by combining high-throughput sequencing with genetic and physical maps⁶⁻⁸. The current assembly spans 675 Mbp and contains about 86% of the contigs ordered and orientated along 24 chromosomes, representing one of the highest quality fish genomes available. Most sea bass chromosomes are highly collinear with those of upper teleosts, confirming the evolutionary stability of fish genomes⁹. We annotate 26,719 genes and identify gene family expansions, some of which are promising candidates for euryhaline adaptation. In addition, we characterize genome-wide variation patterns by RAD-sequencing about 2.5% of the genome in 100 individuals

from the Mediterranean and Atlantic lineages, using three individuals of the congeneric *D. punctatus* as an outgroup. We find that genomic variation within and between sea bass lineages is influenced by large-scale variation in local recombination rates and by diversifying selection translating into variable rates of introgression across the genome following post-glacial secondary contact. These results provide new insights into the genetic correlates of speciation, as it occurs in a typical marine species with high levels of gene flow and a large effective population size.

Results

Genome sequencing and assembly

We sequenced the whole genome of a single meiogynogenetic male of the European sea bass to an average coverage depth of 30x using a combination of whole genome shotgun, mate pair and BAC end sequencing (**Supplementary Table 1**). The reads from three independent sequencing technologies (23.3 Gb) were assembled into a 675 Mb draft genome, with an N50 contig length of 53 kb and an N50 scaffold length of 5.1 Mb (**Supplementary Table 2**). Using a reference radiation hybrid map⁶, linkage maps^{10,11} and collinearity with the genome of *Gasterosteus aculeatus*¹², we assigned 86% of contig sequences (575 Mb) to 24 chromosomal groups⁶ (**Supplementary Table 2**). Comparison between a previous assembly of three chromosomes⁸ and the current assembly resulting from whole genome sequencing (WGS) revealed high similarities with locally improved scaffold orientation (**Supplementary Fig. 1**).

Genome annotation

A combination of *ab initio* gene prediction, homology search and transcript mapping resulted in 26,719 annotated genes (**Supplementary Table 3 and 4**). The 18,253 bp of the whole mitogenome was assembled and annotated, and ranked among the largest in teleost fish (**Supplementary Table 5**).

Repetitive DNA sequences within the sea bass genome accounted for 21.47% of the assembly, with 3.87% of class I transposable elements and 4.19% of DNA transposons (**Supplementary Table 6**), in line with what has been found in other fish genomes^{9,13}. Similarly, the average GC content of the sea bass genome (40.4%) was

comparable to that of other teleosts¹⁴. The distribution of GC content was rather homogeneous within chromosomes (**Supplementary Fig. 2**). We found a lower percentage of GC in non-coding regions (39.6%) compared to protein-coding regions (52.6%), and the third-codon positions GC content was 60.7%, suggesting the effect of selection for codon usage, or a stronger effect of biased gene conversion in coding regions relative to non-coding regions¹⁵. We found almost complete synteny and large blocks of collinearity between sea bass chromosomes and homologous chromosomes of the three most closely related teleost taxa that have a chromosome-scale assembly (*G. aculeatus*, *Oreochromis niloticus* and *Tetraodon nigroviridis*) (**Figure 1, and Supplementary Table 7**). This matches well with the position of sea bass relative to other derived teleosts on a phylogenetic tree reconstructed using 621 1:1 orthologous proteins from 20 sequenced fish genomes (**Figure 2**). Our phylogenomic analysis also supports the recently resolved relationships among fully sequenced percomorph fishes, which group sea basses and sticklebacks in a common clade^{16,17}.

Gene duplication and the evolution of euryhalinity

We evaluated the evolutionary consequences of gene expansion and differential duplicate gene loss on biological function pathways. Some pathways were enriched within certain chromosomes, including neuroactive ligand-receptor interaction (LG2), cell adhesion molecules (CAMs) (LG13 and LG14), endocytosis (LG19), DNA repair and nucleotide excision repair (LG20). Among the enriched pathways, we detected gene family expansions which may have played a role in adaptation to euryhalinity. These include claudins^{18,19}, aquaporins²⁰, arginine-vasotocin (AVT) receptors²¹, prolactin (*PRL*)²² and its receptor (*PRLR*). Although euryhalinity is an old innovation in aquatic vertebrates which has potentially favoured species diversification²³, the sea bass genome provides new insights into the genomic modifications at play in this key adaptation.

While mammalian genomes have 27 claudins²⁴, *D. labrax* has 61 claudin copies (**Supplementary Fig. 3**), thus exceeding the 54 copies found in the zebrafish (*Danio rerio*)²⁵ which are largely explained by the teleost specific whole genome duplication (TGD)²⁶. Chromosome blocks containing claudin genes are duplicated in LG13 (15 genes) and LG14 (14 genes) (**Figure 1 and Figure 3**) and the two claudin clusters of

11 (LG13) and 10 (LG14) tandem genes exhibit conserved synteny to claudins 3 and 4 in human chromosome 7. Expansion of these clusters occurred prior to the TGD event²⁶, since the spotted gar *Lepisosteus oculatus* has both a single synteny block in LG22 (as in humans) and a tandem cluster of 10 claudins (**Figure 3**). However, the coelacanth *Latimeria chalumnae* has a synteny block with only 2 claudin genes, indicating an expansion specific to the Actinopterygii lineage occurred²⁷. While synteny blocks 1 to 3 seem to be part of an ancestral chromosome, both Sarcopterygii and Actinopterygii have gained lineage specific insertions (blocks 4M and 4F) between blocks 1 and 2 (**Figure 3**), which appear to have driven claudin expansion both in mammals and teleosts. After the TGD, the teleost specific duplicate, block 4F on LG14, was translocated between blocks 2 and 3. The LG13 duplicate, block 4F, was eliminated while block 3 was translocated to a distant part of the chromosome, blocks 1 and 2 being maintained together.

The sea bass and zebrafish genomes contain 18 members of the aquaporin family (**Supplementary Fig. 4**), representing the largest repertoire of functional aquaporins in vertebrates²⁰. Both sea bass and zebrafish share duplications of *AQP0a-b*, *AQP1a-b*, *AQP10a-b* and the putative loss of one of the two *AQP4*, *AQP7* and *AQP12* paralogs. However, the *AQP5/1* and one of the *AQP9* paralogs present in *D. rerio* are not found in the sea bass. Conversely, sea bass is the only teleost which retains four *AQP8* copies as a result of TGD²⁸.

Teleost genomes contain a large diversity of *AVT* receptors²¹, with up to 6 in *G. aculeatus*, *O. niloticus* and *O. latipes* and 7 in the sea bass. The latter included duplicated *V1A* and expansion of *V2* from 4 to 5 copies, including a novel *V2B-like* receptor (**Supplementary Fig. 5**).

While the majority of teleosts have maintained single copies of *PRL* and *PRL-like* genes, sea bass and *Takifugu rubripes*²⁹ retained two *PRL-like* genes and *Oryzias latipes* retained two *PRL* genes (**Supplementary Fig. 6**). The *PRL* receptor gene (*PRLR*) is also duplicated in most fish species, but an extra copy of *PRLR-like* homolog is present in sea bass, *T. rubripes*, *G. aculeatus* and in some cichlids (**Supplementary Fig. 7**).

The sea bass genome has the highest number of gene copies linked to ion and water regulation (94 genes) among fully sequenced teleosts. To test the hypothesis that this gene expansion is associated with the degree of euryhalinity in teleost

fishes, a comparison was made to the copy number in the genomes of euryhaline (*T. nigroviridis*, *O. latipes*, *O. niloticus* and *G. aculeatus*) and stenohaline fishes (*D. rerio*, *Gadus morhua* and *T. rubripes*), but no clear pattern was found (**Figure 2**, **Supplementary Table 8**). However, the analysis of the protein-coding sequences of each gene family showed accelerated evolution of the sea bass *PRLR* copies 1-3 and of the novel vasotocin receptor *V2B-L*, suggesting that relaxed purifying selection or positive selection has occurred for these genes in the lineage leading to sea bass (**Supplementary Table 9-12**). Other osmoregulation-related gene duplicates also displayed signatures of positive selection, although not exclusively in sea bass (e.g. *PRL-L2*). Altogether, these results suggest that the lineage leading to sea bass has retained a large number of genes involved in ion and volume regulation since TGD, and this may have facilitated the evolution of some genes conferring high tolerance to rapid salinity changes.

Searches for recently duplicated genes identified six paralogs that are specific to sea bass, five of which exhibit signatures of positive selection at several amino-acid positions in at least one of the duplicates (**Supplementary Table 13**). For one of these genes, the nuclear receptor coactivator 5 (*NCOA5*), sex-differential expression has been reported in Nile tilapia during larval development³⁰. Because sex-determining loci in teleosts often involve recently duplicated genes³¹, *NCOA5* represents an interesting candidate component of the genetic system for sex determination in sea bass.

Genome-wide patterns of polymorphism and recombination

A genetic variation map was produced from 100 wild-caught sea bass from the Atlantic Ocean and western Mediterranean Sea. We additionally analysed three spotted sea bass (*D. punctatus*) individuals as outgroup, in order to identify divergent sites between *D. labrax* and *D. punctatus* and to polarize SNPs within *D. labrax* (**Supplementary Table 14**). Genomic variation within and among individuals, lineages and species was examined at about 178,000 restriction-site-associated DNA (RAD) tags³², providing an average marker density of one pair of RAD-tags every 7.5 kb and a 2.5% genome coverage. A total of 234,148 single nucleotide polymorphisms were found (mean read depth 48x per individual), revealing an average nucleotide diversity (π) of 2.52×10^{-3} in the Atlantic lineage, 2.60×10^{-3} in the

Mediterranean lineage and 2.31×10^{-3} in the spotted sea bass. There was broad variation in nucleotide diversity within chromosomes, with a frequently observed five-fold reduced diversity in central chromosomal regions compared to the chromosomal extremities (**Figure 4A**). Polymorphism was negatively correlated with divergence between *D. labrax* and the outgroup species *D. punctatus* ($R^2=0.085$, $p<10^{-15}$), indicating that the genomic landscape of relative divergence between species was slightly influenced by chromosomal variations in diversity within *D. labrax* (**Figure 4B**). Recombination rates were markedly reduced in central chromosomal regions (**Figure 4C**) and were positively correlated to nucleotide diversity ($R^2=0.586$, $p<10^{-15}$), while the correlation was negative between recombination and divergence ($R^2=0.078$, $p<10^{-15}$). Thus, the possible mutagenic role of recombination could not explain the patterns of nucleotide diversity within the sea bass genome, providing evidence for Hill-Robertson effects (including background selection and/or hitchhiking effects)^{33,34}. Consistent with the effects of selection on SNP variation at a local genomic scale, there was a lower mean expected nucleotide diversity in exonic ($H_e = 0.067$) compared to intronic ($H_e = 0.075$) and intergenic ($H_e = 0.073$) regions. We also observed a diminishing average heterozygosity with decreasing distance to the nearest exon on a very local scale (<200 bp) (**Supplementary Fig. 8**), reflecting either Hill-Robertson effects or the direct impact of purifying selection on evolutionarily constrained sites in both coding and cis-regulatory sequences.

Genomic landscape of differentiation between lineages

Genetic distinctiveness between the Atlantic and Mediterranean lineages of *D. labrax* was revealed by Principal Component Analysis (PCA), in which both lineages appeared equally distant from the closely related spotted sea bass species (**Figure 5A**). The genome-wide average genetic differentiation between lineages was low ($F_{ST} = 0.028$) and consistent with previous estimates based on microsatellite markers^{4,35,36}. However, SNP-by-SNP F_{ST} estimates between lineages were heterogeneously distributed across the genome, with highly differentiated markers (some of which reached differential fixation) usually clustering within regions of several hundred Kb to >1Mb. These genomic islands of differentiation were present in all chromosomes except for LG24 and tended to map disproportionately to central chromosomal regions (**Figure 4D**). The local averaged F_{ST} was negatively correlated

with the local recombination rate ($R^2=0.257$, $p<10^{-15}$) and nucleotide diversity ($R^2=0.207$, $p<10^{-15}$), as previously described in other species^{37,38}. However, a few highly differentiated regions were also found outside low-recombining and poorly polymorphic chromosomal regions (e.g. LG13). Therefore, the heterogeneous genomic landscape of differentiation between lineages was only partially explained by the reduced diversity found in the regions of reduced recombination.

Demographic divergence history

To determine whether and when genomic differentiation patterns were influenced by gene flow, the past demography of European sea bass lineages was inferred using a composite likelihood approach³⁹. Seven alternative models of historical divergence were fitted to the joint allele-frequency spectrum (AFS) of Atlantic and Mediterranean lineages (**Figure 5B**), including scenarios of strict isolation, isolation-with-migration, ancient migration and secondary contact. The observed AFS was not satisfactorily reproduced by standard demographic models that assume the gene flow parameter to be shared among loci (**Supplementary Fig. 9**). A custom secondary contact model with heterogeneous gene flow across the genome⁴⁰ (**Figure 5C**) produced a significantly higher fit compared to all other alternative models (**Figure 5D**). This analysis revealed several key features of the divergence history of sea bass lineages (**Supplementary Table 15**). Divergence accumulated during a period of isolation at least 20 times longer than the age of secondary contact. Genetic introgression following secondary contact was strongly asymmetric from the Atlantic to Mediterranean lineage, and occurred at highly variable rates across the genome. We estimated that about 35% of the genome did not freely introgress from one lineage to the other, which mirrors the proportion of the genome in which the windowed F_{ST} lies above the genome-wide average differentiation (**Figure 4D**). Thus, the genomic islands of differentiation between sea bass lineages appear to have resulted from the erosion of divergence through differential introgression.

Discussion

The European sea bass genome assembly represents one of the first high-quality draft genomes available for an aquaculture fish species and will provide a valuable resource for future evolutionary analyses and genetic improvement. Compared to

solely NGS-based genome assemblies, the sea bass genome assembly has benefited from low-coverage Sanger sequencing reads, which combined with data from two different NGS platforms, led to increased contig size. Scaffolding was also improved by combining mate-pairs with BAC-end sequences, and a chromosome-scale assembly was obtained using available genetic linkage and radiation hybrid maps anchoring scaffolds to the 24 chromosomes comprising the sea bass karyotype. This genome assembly is a prerequisite to whole-genome or targeted resequencing, RNA-seq, ChIP-seq and Methyl-seq experiments, and will therefore catalyze genomic studies in sea bass. Additionally, it will help to improve the assembly of newly sequenced genomes from related orders, since large regions of collinearity have been observed with the genome assemblies of other teleost fishes. The sea bass genome sequence may thus facilitate genomic studies in other economically important fishes.

Expansions of gene families related to ion and volume regulation were likely mediated through an elevated rate of duplicate gene retention after TGD, providing a genetic basis for adaptation to euryhalinity in the European sea bass and possibly other highly adaptable species. This finding questions the relative importance of whole genome duplication events versus single gene duplication in the evolution of euryhalinity in fish, which should motivate broader scale comparative genomic and experimental studies in the future.

The European sea bass also provides an interesting model for understanding the evolutionary mechanisms involved in speciation. Our results imply genetically-based reproductive barriers reducing gene flow between Atlantic and Mediterranean sea bass lineages. We reveal important aspects of the genetic architecture of differentiation between these two lineages, such as the disproportionate mapping of the genomic islands of differentiation within low-recombining regions. Evidence for differential gene flow after secondary contact implies that the genomic islands in sea bass are not simply an incidental consequence of reduced diversity in low-recombining regions⁴¹, but also the product of reduced gene flow in these regions. Furthermore, some genomic islands of high differentiation are also observed in regions of high diversity. Therefore, the genomic landscape of differentiation in sea bass likely results from the preferential erosion of divergence in the genomic regions with high recombination rates. The question of the timing of gene flow has been

partly answered by our modelling approach, which dated the secondary contact to ca. 11,500 years BP, and estimated the divergence time to ca. 270,000 years BP (using a per-generation mutation rate of 10^{-8} per bp and a generation time of 5 years). This puts the onset of secondary gene flow to the last glacial retreat and strongly supports the role of distributional range shifts caused by Pleistocene glacial periods in promoting divergence. The true evolutionary history is admittedly more complex, since quaternary glacial oscillations may have resulted in intermittent gene flow during divergence. This may explain the contrast between the divergence captured by our analysis of the nuclear genome and the 2.8% divergence found in the non-recombining mitochondrial genome, which may be involved in old cyto-nuclear incompatibilities that have remained protected from fixation for a much longer period than the recent history revealed by the coalescence of the rest of the genome. This new picture of the dual origin of the genetic variability in European sea bass brings important data to explain the long-standing observation of genetic discontinuities at the Almeria-Oran front in many marine species⁴². It reinforces the hypothesis of anciently diverged genetic backgrounds being trapped by the Atlantic/Mediterranean environmental boundary⁴³, and further supports the role of allopatric isolation as the main driver of marine speciation.

Methods

Sequencing strategy

All sequencing libraries were constructed from the genomic DNA of a single Adriatic sea bass individual (meiogynogenetic male 57)^{44,45} (kindly provided by the late A. Libertini, CNR, Venice, Italy through J. B. Taggart, University of Stirling, Stirling, UK). The BAC library⁴⁴ was obtained from the German Resources Center for Genome Research (RZPD, Berlin, Germany). Plasmid libraries were constructed as previously described⁷. Template preparation was performed automatically at the Max Planck Institute for Molecular Genetics (MPIMG, Berlin, Germany). Purification was based on size-selective precipitation in polyethylene-glycol 6000/2-propanol mixtures⁴⁶. Template DNA was sequenced on ABI3730xl capillary sequencers and raw sequencing data was processed using the PHRED basecaller⁴⁷ and LUCY⁴⁸ for quality clipping and vector-clipping.

WGS libraries that were submitted to pyrosequencing were constructed according to the manufacturers' protocols (Roche) and sequenced on a ROCHE 454 FLX Titanium sequencer. Eight mate-pair libraries of approximately 20 kb insert size were constructed at 454 Life Sciences (Branford, USA) to support genome scaffolding. For each library, half of a picotiter plate was sequenced on the GS FLX Titanium system. A small-insert library (300 bp) for Illumina sequencing was constructed using the NEBNext DNA Sample Prep Reagent Set 1 (NEB, Ipswich, USA). For the two larger insert libraries (500 bp and 1000 bp), only 5 PCR cycles were performed to reduce amplification bias effects. All libraries were sequenced with 100 bp paired-end runs on an Illumina GAIIx instrument.

Genome assembly

Before assembly, Illumina sequencing reads were end-clipped to keep the largest part of a read (> 64 bp) in which no quality value below 11 occurred. Duplicate reads and residual sequencing adapters were removed. 454 reads from 20 kb mate pair libraries were filtered for duplicates and chimeric mates by mapping them to older assembly versions of the *D. labrax* genome using gsMapper v2.3 (Roche) and confirming collinearity with *G. aculeatus* and *O. latipes*. Sanger reads and 454 single-end reads were directly used for assembly using Celera Assembler v6.1 (CA6.1)⁴⁹. Several softwares for WGS assembly were tested, among which SOAPdenovo (for Illumina data), Newbler (Roche) and CA6.1. CA6.1 was the best performing, although the first assemblies needed editing due to inconsistencies in the long range mate pair data. These issues were solved by removing duplicates and screening for potential chimeric mate pairs by making use of the strong collinearity of sea bass with the stickleback¹². After WGS assembly, long range continuity was improved by re-scaffolding the CA6.1 scaffolds using the BAMBUS scaffolder⁵⁰, including the 454 20 kb mate pairs and BAC-ends. These scaffolds were mapped against the *G. aculeatus* and *O. latipes* genomes by Megablast. Scaffolds were grouped into sets that were assigned to distinct chromosomal groups in these species to reduce the complexity of chromosome assembly.

Finally, markers from the radiation hybrid map⁶, genetic linkage maps^{10,11}, BAC-end and 454 20 kb mate pairs were mapped onto these scaffold groups and CONSED⁵¹ was applied to manually build chromosomal sized superscaffolds.

Chromosome names conformed with the naming adopted in the radiation hybrid map. Sequences that could not be ordered into chromosomal groups were concatenated into a superscaffold of unordered genomic pieces (UN).

Transcriptome sequencing and assembly

Paired-end RNA-sequencing was performed using sea bass offspring obtained by crossing wild parents at the Ifremer Station Expérimentale d'Aquaculture (Palavas-les-Flots, France). Four pools of 3-4 individual liver and intestine total RNA extracts were assembled to prepare 4 RNAseq libraries of distinct Mediterranean and Atlantic origin (**Supplementary Table 4**) at the CERBM (Université de Strasbourg, France). mRNA was purified from total RNA using poly-T oligo-attached magnetic beads and fragmented using divalent cations at 95°C during 5 min. After reverse transcription with random primers and ligation to adapters, enrichment was performed through 13 cycles of PCR amplification. PCR products were then purified with AMPure beads (Agencourt). Size selection of ~250-350 bp fragments was performed by electrophoresis on a 2% agarose gel. Libraries were sequenced on 4 lanes of an Illumina GAllx using 2x100 pb paired-end reads. After library cleaning with Trimmomatic⁵², reads from each of the 4 lanes were mapped independently on the draft sequence of the sea bass genome with Tophat and then assembled with Cufflinks⁵³, and a reference transcriptome was built by merging results of the four lanes with Cuffmerge.

Genome annotation

The *ab initio* gene prediction was done with GENESCAN⁵⁴. For homology based predictions, we first mapped known teleost proteins from Ensembl and GenBank databases onto the sea bass genome using SPALN aligner v2.12⁵⁵. Custom scripts were used to choose the best scoring protein match in a cluster of matches defined by exact exon-exon matches in a first iteration and overlapping exons in a second iteration. Secondly, the CDS models from SPALN were combined with assembled RNAseq data by splitting the gtf files of all predictions and transcripts into overlapping exon-intron-exon fragments and merging them using Cuffmerge⁵³. This resulted in a high number of possible transcript models. We used Transdecoder [<http://transdecoder.sourceforge.net>] to assign coding sequences to these models

and used the length of the CDS as a score for each model. We weighted the scores of these different models based on their origin (highest rank: RNAseq only, lowest rank: SPALN only). Afterwards, the best scoring transcript from a cluster of redundant CDS was selected as the reference gene model.

Reference genes resulting from our combined annotation were functionally annotated using Blast2GO v2.6.4⁵⁶ and a custom blast database consisting of vertebrate proteins from GenBank. Additionally, the gene model proteins were submitted to InterProScan to identify shared signatures with proteins of known function. Functional annotation of RNAseq data was performed using blastx against Ensembl fish species proteomes and the NCBI *nr* database. Results were contrasted with annotation based on the direct homology-based method, and permitted annotation of new genes and correction of some gene predictions.

Whole genome alignments

Reference genome sequences of seven teleost fishes (*G. aculeatus*, *O. niloticus*, *T. nigroviridis*, *T. rubripes*, *O. latipes*, *G. morhua*, *D. rerio*) were downloaded from the Ensembl website. The LAST tool was applied to produce fast and sensitive whole genome alignments. We screened for best one to one alignments using custom scripts removing repetitive and suboptimal alignments. Blocks of shared collinearity were constructed using the BlockDisplaySatsuma script from the Satsuma v1.17 package⁵⁷. Blocks of collinearity were further combined allowing unaligned regions of up to 100 kb and visualized in the sea bass genome browser.

Phylogenetic reconstruction

Twenty publicly available genomes were downloaded and formatted as SPALN v2.12 databases. The Ensembl proteins from *O. niloticus* were used to predict genes in all species by homology using SPALN v2.12. Predicted proteins for each species were filtered to retain only the best scoring hit to *O. niloticus*, showing nearly complete coverage and a minimum alignment of 60% of amino acid residues to an *O. niloticus* protein. We only retained 621 1:1 orthologous proteins that were found in all 20 genomes. Multiple alignments computed for each protein using ClustalW2.1 were then concatenated and Prottest used to determine the best amino acid substitution model (JTT I+G) and phyML used to build phylogenetic trees. Other phylogenetic

reconstruction methods were also tested (Neighbor Joining, FastTreeMP, MrBayes) and in all trees the nodes were highly supported and only a single difference in tree topology appeared between NJ and the other methods. Visualization of the trees was done in TreeDyn.

Enrichment pathway analysis

Chromosome gene sets were tested for pathway enrichments in comparison to the whole genome gene set. Sea bass genes were queried against KEGG orthology (KO) database using KOBAS⁵⁸. KO terms of the annotated genes within each sea bass chromosome were used to identify statistically enriched related pathways in chromosomal regions after applying the FDR correction.

Detection of gene synteny and collinearity

MCScanX's algorithm was used for perform synteny and collinearity detection across the genome of different species, as well as within the sea bass genome to search for collinear blocks indicative of duplications events and gene family expansions. We used the fraction of collinear genes as a metric to identify collinear blocks. Circular genome representations were created using Circos⁵⁹.

Detection of duplicated genes

Sea bass specific gene duplications were detected using a pipeline that interrogates the Ensembl database of orthology groups Compara, using genes from eight teleost genomes to build Compara gene trees (*D. rerio*, *G. morhua*, *O. niloticus*, *T. rubripes*, *T. nigroviridis*, *X. maculatus*, *G. aculeatus*, *O. latypes*). Putative duplicated genes were then analysed in a rigorous phylogenetic context to assess true orthology. Alignment of protein sequences of all putative orthologs and closely related paralogs in all vertebrates was carried out for each group of duplicates. Amino acid sequences were aligned using MAFFT⁶⁰ under E-INS-I or L-INS-I options. Sequences with insufficient information were discarded. Aligned coding sequences were used to reconstruct a gene tree with PHYML 3.0⁶¹ using default options. The tree topology was analysed for compatibility with gene duplication occurring in the sea bass lineage. For confirmed duplicated genes, the amino acid alignment was then transferred into Translator X⁶² to guide nucleotide sequence alignment and obtain

aligned codon sequences. The tree topology was used as user tree option in the analysis of branch- and site-specific codon evolution, which was implemented in the Data Monkey web server (<http://www.datamonkey.org/>). Branch Site REL⁶³, a branch-specific test for positive selection was used to assess the presence of significantly divergent branches in the gene tree. A mixed effects model of evolution⁶⁴ test was used to assess the presence of codons/positions under positive selection.

RAD Sequencing, SNP discovery and genotyping

Wild sea bass individuals were sampled from both the Atlantic Ocean (n = 50) and the western Mediterranean Sea (n = 50), where two distinct lineages have been documented for *D. labrax*⁴ (**Supplementary Table 9**). Three individuals from the closely related spotted sea bass *D. punctatus* were sampled in a Tunisian lagoon and used as an outgroup. Genomic DNA was isolated using the DNeasy Blood and Tissue Kit (Qiagen) and digested with *Bam*HI. Seven RAD libraries were constructed by multiplexing 15 uniquely barcoded individuals per library, following an adaptation of the original protocol³². Each library was subsequently sequenced on a separate lane of an Illumina HiSeq2000 instrument with 101 bp single-end reads.

Illumina reads were demultiplexed and quality filtered using Stacks v1.07⁶⁵. Cleaned individual reads were mapped to the reference genome using Bowtie v2.0 with the very-sensitive option⁶⁶, allowing at most 3 mismatches per alignment. We then called SNPs from the aligned reads using the Stacks v1.07 pipeline, with a minimum read depth of 5x per individual per allele to infer individual genotypes. Only RAD loci that were present in at least 70% of the samples in each population were considered for nucleotide diversity analysis.

Population genetic analyses

Genomic patterns of nucleotide diversity and genetic differentiation were computed in Stacks v1.07 as the weighted average of π and F_{ST} in 150 kb windows⁶⁵. We used a custom script to calculate the mean proportion of fixed differences per bp (d_f) between *D. labrax* and *D. punctatus* in 150 kb windows. SNPs were annotated to exonic, intronic, or intergenic regions using *SNPdat*⁶⁷, and the mean expected heterozygosity in different regions was calculated with *VCFtools* v0.1.11⁶⁸.

Variable recombination rates along each chromosome ($\rho = 4N_e r$ per kb) were estimated using a Bayesian reversible-jump MCMC scheme under the crossing-over model of *interval* in *LDhat*⁶⁹, with 10 million iterations and 5 million burn-ins. Population structure was examined using a PCA in SNPRelate⁷⁰. The demographic history of Atlantic and Mediterranean sea bass populations was inferred from their joint SFS in *δaδi v1.6.3*³⁹. We used 109329 SNPs that were polymorphic in *D. labrax* but fixed in *D. punctatus* to determine the most parsimonious ancestral allele in *D. labrax*. The joint SFS was projected to 45 individuals in each population to avoid missing genotypes. We considered seven alternative models of historical divergence: Strict Isolation (*S*), Isolation-with-Migration (*IM*), Ancient Migration (*AM*), Secondary Contact (*SC*), and their heterogeneous migration rates versions: *IM2m*, *AM2m* and *SC2m* (**Supplementary Fig. 9**).

References

- 1 Pickett, G. D. & Pawson, M. G. *Sea Bass. Biology, Exploitation and Conservation*. Vol. 12 Chapman & Hall, London. (1994).
- 2 Boutet, I., Long Ky, C. L. & Bonhomme, F. A transcriptomic approach of salinity response in the euryhaline teleost, *Dicentrarchus labrax*. *Gene* **379**, 40-50 (2006).
- 3 Ohno, S. *Evolution by gene duplication*. Springer (1970).
- 4 Lemaire, C., Versini, J. J. & Bonhomme, F. Maintenance of genetic differentiation across a transition zone in the sea: discordance between nuclear and cytoplasmic markers. *Journal of Evolutionary Biology* **18**, 70-80 (2005).
- 5 Allendorf, F. W., Hohenlohe, P. A. & Luikart, G. Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11**, 697-709 (2010).
- 6 Guyon, R. *et al.* A radiation hybrid map of the European sea bass (*Dicentrarchus labrax*) based on 1581 markers: Synteny analysis with model fish genomes. *Genomics* **96**, 228-238 (2010).
- 7 Kuhl, H. *et al.* The European sea bass *Dicentrarchus labrax* genome puzzle: comparative BAC-mapping and low coverage shotgun sequencing. *BMC Genomics* **11**, 68 (2010).
- 8 Kuhl, H. *et al.* Directed sequencing and annotation of three *Dicentrarchus labrax* L. chromosomes by applying Sanger- and pyrosequencing technologies on pooled DNA of comparatively mapped BAC clones. *Genomics* **98**, 202-212 (2011).
- 9 Schartl, M. *et al.* The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nat Genet* **45**, 567-572 (2013).
- 10 Chistiakov, D. A. *et al.* A combined AFLP and microsatellite linkage map and pilot comparative genomic analysis of European sea bass *Dicentrarchus labrax* L. *Anim Genet* **39**, 623-634 (2008).
- 11 Chistiakov, D. A. *et al.* A microsatellite linkage map of the European sea bass *Dicentrarchus labrax* L. *Genetics* **170**, 1821-1826 (2005).
- 12 Jones, F. C. *et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55-61 (2012).
- 13 Chen, S. *et al.* Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat Genet* **46**, 253-260 (2014).
- 14 Han, L. & Zhao, Z. Comparative analysis of CpG islands in four fish genomes. *Comparative and functional genomics* **2008** (2008).
- 15 Duret, L. & Galtier, N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annual review of genomics and human genetics* **10**, 285-311 (2009).
- 16 Betancur-R, R. *et al.* The tree of life and a new classification of bony fishes. *PLoS currents* **5** (2013).
- 17 Near, T. J. *et al.* Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proceedings of the National Academy of Sciences* **110**, 12738-12743 (2013).
- 18 Loh, Y. H., Christoffels, A., Brenner, S., Hunziker, W. & Venkatesh, B. Extensive expansion of the claudin gene family in the teleost fish, *Fugu rubripes*. *Genome Res* **14**, 1248-1257 (2004).
- 19 Tipsmark, C. K., Baltzegar, D. A., Ozden, O., Grubb, B. J. & Borski, R. J. Salinity regulates claudin mRNA and protein expression in the teleost gill. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology* **294**, R1004-R1014 (2008).
- 20 Tingaud-Sequeira, A. *et al.* The zebrafish genome encodes the largest vertebrate repertoire of functional aquaporins with dual paralogy and substrate specificities similar to mammals. *BMC Evolutionary Biology* **10**, 38 (2010).

575 21 Ocampo Daza, D., Lewicka, M. & Larhammar, D. The oxytocin/vasopressin receptor family
576 has at least five members in the gnathostome lineage, including two distinct V2 subtypes.
577 *General and Comparative Endocrinology* **175**, 135-143 (2012).

578 22 Manzon, L. A. The role of prolactin in fish osmoregulation: a review. *Gen Comp Endocrinol*
579 **125**, 291-310 (2002).

580 23 Schultz, E. T. & McCormick, S. D. in *Euryhaline Fishes* (eds Stephen D McCormick, AP Farrell,
581 & CJ Brauner) Academic Press, Oxford (2013).

582 24 Mineta, K. *et al.* Predicted expansion of the claudin multigene family. *FEBS letters* **585**, 606-
583 612 (2011).

584 25 Baltzegar, D. A., Reading, B. J., Brune, E. S. & Borski, R. J. Phylogenetic revision of the claudin
585 gene family. *Marine genomics* **11**, 17-26 (2013).

586 26 Amores, A., Catchen, J., Ferrara, A., Fontenot, Q. & Postlethwait, J. H. Genome evolution and
587 meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost
588 genome duplication. *Genetics* **188**, 799-808 (2011).

589 27 Nikaido, M. *et al.* Coelacanth genomes reveal signatures for evolutionary transition from
590 water to land. *Genome research* **23**, 1740-1748 (2013).

591 28 Cerdà, J. & Finn, R. N. Piscine aquaporins: an overview of recent advances. *Journal of*
592 *Experimental Zoology Part A: Ecological Genetics and Physiology* **313A**, 623-650 (2010).

593 29 Wang, Y., Li, J., Yan Kwok, A. H., Ge, W. & Leung, F. C. A novel prolactin-like protein (PRL-L)
594 gene in chickens and zebrafish: cloning and characterization of its tissue expression. *Gen*
595 *Comp Endocrinol* **166**, 200-210 (2010).

596 30 Tao, W. *et al.* Characterization of gonadal transcriptomes from Nile tilapia (*Oreochromis*
597 *niloticus*) reveals differentially expressed genes. *PLoS ONE* **8**, e63604.
598 doi:63610.61371/journal.pone.0063604 (2013).

599 31 Kikuchi, K. & Hamaguchi, S. Novel sex-determining genes in fish and sex
600 chromosome evolution. *Developmental Dynamics* **242**, 339-353 (2013).

601 32 Baird, N. A. *et al.* Rapid SNP discovery and genetic mapping using sequenced RAD markers.
602 *PLoS ONE* **3**, e3376 (2008).

603 33 Charlesworth, B., Morgan, M. & Charlesworth, D. The effect of deleterious mutations on
604 neutral molecular variation. *Genetics* **134**, 1289-1303 (1993).

605 34 Maynard Smith, J. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet Res* **23**, 23-
606 35 (1974).

607 35 Quéré, N. *et al.* Gene flow at major transitional areas in sea bass (*Dicentrarchus labrax*) and
608 the possible emergence of a hybrid swarm. *Ecology and Evolution* **2**, 3061-3078 (2012).

609 36 Naciri, M., Lemaire, C., Borsa, P. & Bonhomme, F. Genetic study of the
610 Atlantic/Mediterranean transition in sea bass (*Dicentrarchus labrax*). *Journal of Heredity* **90**,
611 591-596 (1999).

612 37 Roesti, M., Hendry, A. P., Salzburger, W. & Berner, D. Genome divergence during
613 evolutionary diversification as revealed in replicate lake–stream stickleback population pairs.
614 *Molecular Ecology* **21**, 2852-2862 (2012).

615 38 Nachman, M. W. & Payseur, B. A. Recombination rate variation and speciation: theoretical
616 predictions and empirical results from rabbits and mice. *Philosophical Transactions of the*
617 *Royal Society B: Biological Sciences* **367**, 409-421 (2012).

618 39 Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the Joint
619 Demographic History of Multiple Populations from Multidimensional SNP Frequency Data.
620 *PLoS Genet* **5**, e1000695 (2009).

621 40 Roux, C., Tsagkogeorga, G., Bierne, N. & Galtier, N. Crossing the species barrier: genomic
622 hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular*
623 *biology and evolution* **30**, 1574-1587 (2013).

624 41 Cruickshank, T. E. & Hahn, M. W. Reanalysis suggests that genomic islands of speciation are
625 due to reduced diversity, not reduced gene flow. *Molecular Ecology* **23**, 3133-3157 (2014).
626 42 Patarnello, T., Volckaert, F. A. M. J. & Castilho, R. Pillars of Hercules: is the Atlantic–
627 Mediterranean transition a phylogeographical break? *Molecular Ecology* **16**, 4426-4444
628 (2007).
629 43 Bierne, N., Welch, J., Loire, E., Bonhomme, F. & David, P. The coupling hypothesis: why
630 genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**, 2044-2072
631 (2011).
632 44 Whitaker, H. A., McAndrew, B. J. & Taggart, J. B. Construction and characterization of a BAC
633 library for the European sea bass *Dicentrarchus labrax*. *Animal Genetics* **37**, 526 (2006).
634 45 Francescon, A. *et al.* Assessment of homozygosity and fertility in meiotic gynogens of the
635 European sea bass (*Dicentrarchus labrax* L.). *Aquaculture* **243**, 93-102 (2005).
636 46 Kuhl, H. *et al.* A Comparative BAC map for the gilthead sea bream (*Sparus aurata* L.). *Journal*
637 *of Biomedicine and Biotechnology* **2011**, 1-7 (2011).
638 47 Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error
639 probabilities. *Genome Res* **8**, 186-194 (1998).
640 48 Li, S. & Chou, H. H. LUCY2: an interactive DNA sequence quality trimming and vector removal
641 tool. *Bioinformatics* **20**, 2865-2866 (2004).
642 49 Miller, J. R. *et al.* Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics*
643 **24**, 2818-2824 (2008).
644 50 Pop, M., Kosack, D. S. & Salzberg, S. L. Hierarchical scaffolding with Bambus. *Genome Res* **14**,
645 149-159 (2004).
646 51 Gordon, D. & Green, P. Consed: a graphical editor for next-generation sequencing.
647 *Bioinformatics* **29**, 2936-2937 (2013).
648 52 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence
649 data. *Bioinformatics*, btu170 (2014).
650 53 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments
651 with TopHat and Cufflinks. *Nat Protoc* **7**, 562-578 (2012).
652 54 Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *Journal*
653 *of Molecular Biology* **268**, 78-94 (1997).
654 55 Iwata, H. & Gotoh, O. Benchmarking spliced alignment programs including Spaln2, an
655 extended version of Spaln that incorporates additional species-specific features. *Nucleic*
656 *Acids Res* **40**, e161 (2012).
657 56 Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in
658 functional genomics research. *Bioinformatics* **21**, 3674-3676 (2005).
659 57 Grabherr, M. G. *et al.* Genome-wide synteny through highly sensitive sequence alignment:
660 Satsuma. *Bioinformatics* **26**, 1145-1151 (2010).
661 58 Wu, J., Mao, X., Cai, T., Luo, J. & Wei, L. KOBAS server: a web-based platform for automated
662 annotation and pathway identification. *Nucleic Acids Research* **34**, W720-W724 (2006).
663 59 Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome*
664 *Research* **19**, 1639-1645 (2009).
665 60 Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7:
666 Improvements in Performance and Usability. *Molecular Biology & Evolution* **30**, 772-780
667 (2013).
668 61 Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood
669 phylogenies: assessing the performance of PhyML 3.0. *System Biology* **59**, 307-321 (2010).
670 62 Abascal, F., Zardoya, R. & Telford, M. J. TranslatorX: multiple alignment of nucleotide
671 sequences guided by amino acid translations. *Nucleic Acids Research* **38** (2010).
672 63 Kosakovsky, P. S. L. *et al.* A random effects branch-site model for detecting episodic
673 diversifying selection. *Molecular Biology & Evolution* **28**, 3033-3043 (2011).

674 64 Murrell, B. *et al.* Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLoS*
675 *Genetics* **8**, e1002764. doi:1002710.1001371/journal.pgen.1002764 (2012).
676 65 Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. Stacks: Building
677 and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* **1**,
678 171-182 (2011).
679 66 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**,
680 357-359 (2012).
681 67 Doran, A. G. & Creevey, C. J. Snpdat: Easy and rapid annotation of results from de novo snp
682 discovery projects for model and non-model organisms. *BMC bioinformatics* **14**, 45 (2013).
683 68 Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011).
684 69 McVean, G. A. *et al.* The fine-scale structure of recombination rate variation in the human
685 genome. *Science* **304**, 581-584 (2004).
686 70 Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal
687 component analysis of SNP data. *Bioinformatics* **28**, 3326-3328 (2012).
688
689

Accession numbers. UCSC genome browser of the sea bass genome: <http://sea.bass.mpipz.mpg.de>. Sequencing reads have been deposited under the NCBI Sequence Read Archive (Contigs: CBXY010000001-CBXY010037781; chromosomal scale assembly: HG916827-HG916851). GenBank accession numbers for mitochondrial genomes (*D. labrax* Mediterranean: KJ168065; *D. labrax* Atlantic: KJ168064; *D. punctatus*: KJ168066).

ACKNOWLEDGMENTS This work was initiated during the MARINE GENOMICS EUROPE project EU-FP6 505403 and supported by grants from the Max Planck Society (MPG for genome sequencing, assembly and annotation), LIFECYCLE EU-FP7 222719 and BMBF-01GS0805 to RR, and the French ANR grants LABRAD-SEQ 11-PDOC-009-01 to PAG and REGULBASS 09-GENM-003 to BG. MT received a fellowship from the MPG, HK was partly financed by LIFECYCLE EU-FP7 222719 and FK by the BMBF-01GS0805 grant. LB was supported by the grant Progetto 12/MI/2004 from the Veneto Region. RSTM and BL were in receipt of fellowships SFRH/BPD/66742/2009 and SFRH/BPD/89889/2012 from the Foundation for Science and Technology of Portugal.

AUTHOR CONTRIBUTIONS MT, HK and PAG contributed equally to this work. PAG, MT, HK, AVMC wrote the manuscript. FB, AVMC, RR and FAMV initiated the project. RR organized financial support and genome sequencing. PAG, MT and FB coordinated manuscript writing. FAMV and FP provided scientific input. HK, JH and FK carried out genome sequencing. HK produced the genome assembly, alignments and browser content. ED, BG produced RNAseq data. Gene annotation was performed by HK, MT, ED and KB. BL, RSTM, DMP and AVMC carried out genome pathway enrichment, synteny and gene expansion analysis. ED and HK performed the phylogenetic analyses. LB performed sea bass-specific gene duplication analyses. PAG, FB and NB produced and analysed RAD sequencing data and performed historical demographic analysis. SK, KS and RD maintained software packages, IT infrastructures and organized sequence submission. RR, FB and AVMC supported the project as senior authors.

722 **COMPETING FINANCIAL INTERESTS** The authors declare no competing financial
723 interests.
724

Figures

Figure 1: Collinear blocks showing the overall degree of synteny between the European sea bass (*Dicentrarchus labrax*) genome and seven other publicly available teleost genomes represented outside the sea bass chromosomal ring. From the inner to the outer layer: *G. aculeatus*, *O. latipes*, *T. nigroviridis*, *D. rerio*, *O. niloticus*, *T. rubripes* and *G. morhua*. Sea bass chromosomes (LGn) show conserved synteny with the assemblies of *G. aculeatus*, *O. latipes*, *T. nigroviridis* and *D. rerio*, while *O. niloticus*, *T. rubripes* and *G. morhua* are still scattered into many ungrouped scaffolds as reflected by tracks of different colours along the chromosomes. The colour code is species specific. Blocks of collinearity between sea bass chromosomes are represented by grey inner links. Red inner links represent blocks of collinearity containing claudin genes.

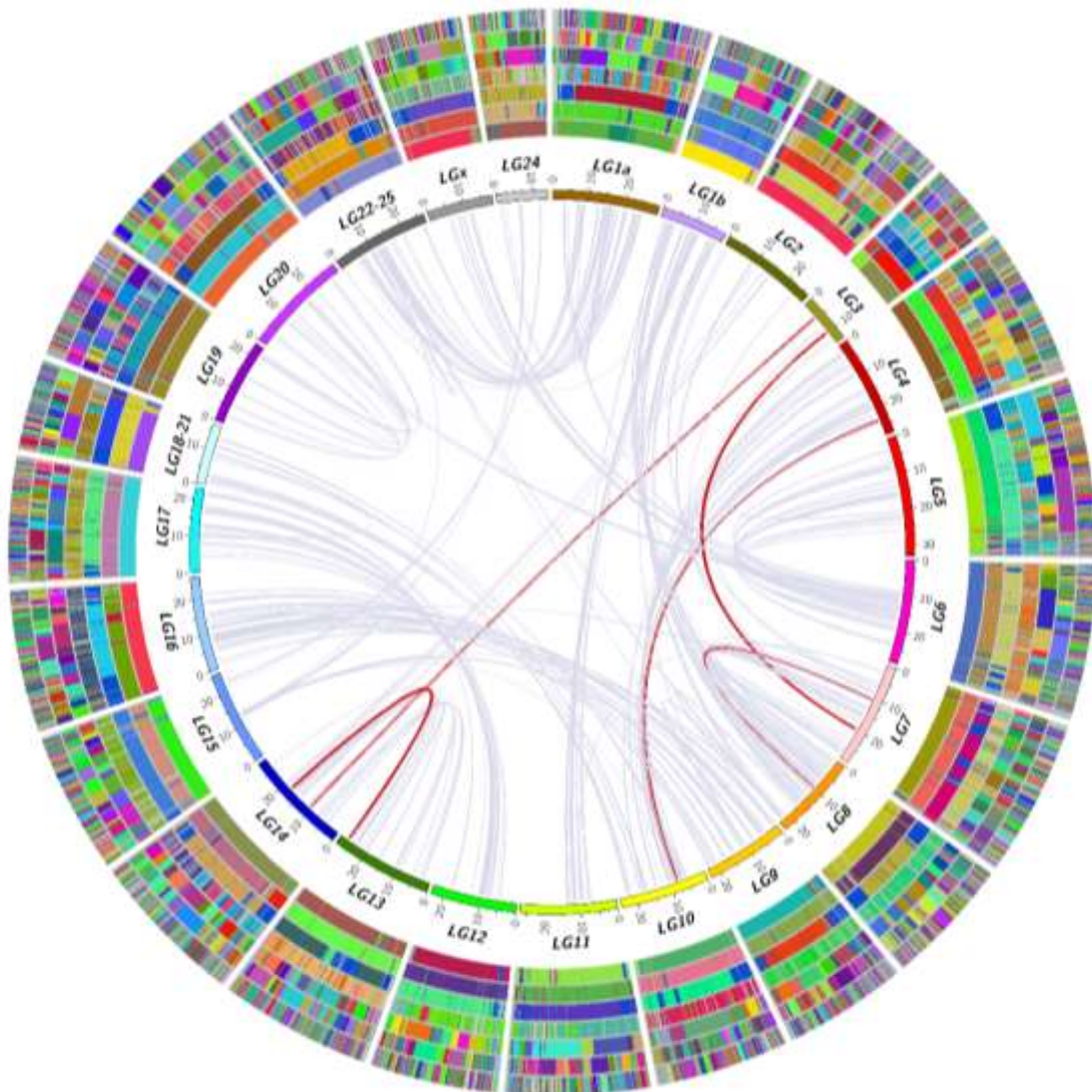


Figure 2: Phylogenetic tree based on 621 1:1 high-quality orthologous protein-coding genes from 20 sequenced fish genomes, showing the relationships between European sea bass (*Dicentrarchus labrax*) and other fish species (half of which belong to the Series Percomorpha). Node support is indicated by bootstrap values. The lancelet (*Branchiostoma floridae*) was used as the outgroup. The shaded box indicates the total number of gene copies across the five expanded gene families involved in euryhalinity for highly euryhaline (red), euryhaline (green) and stenohaline (blue) species.

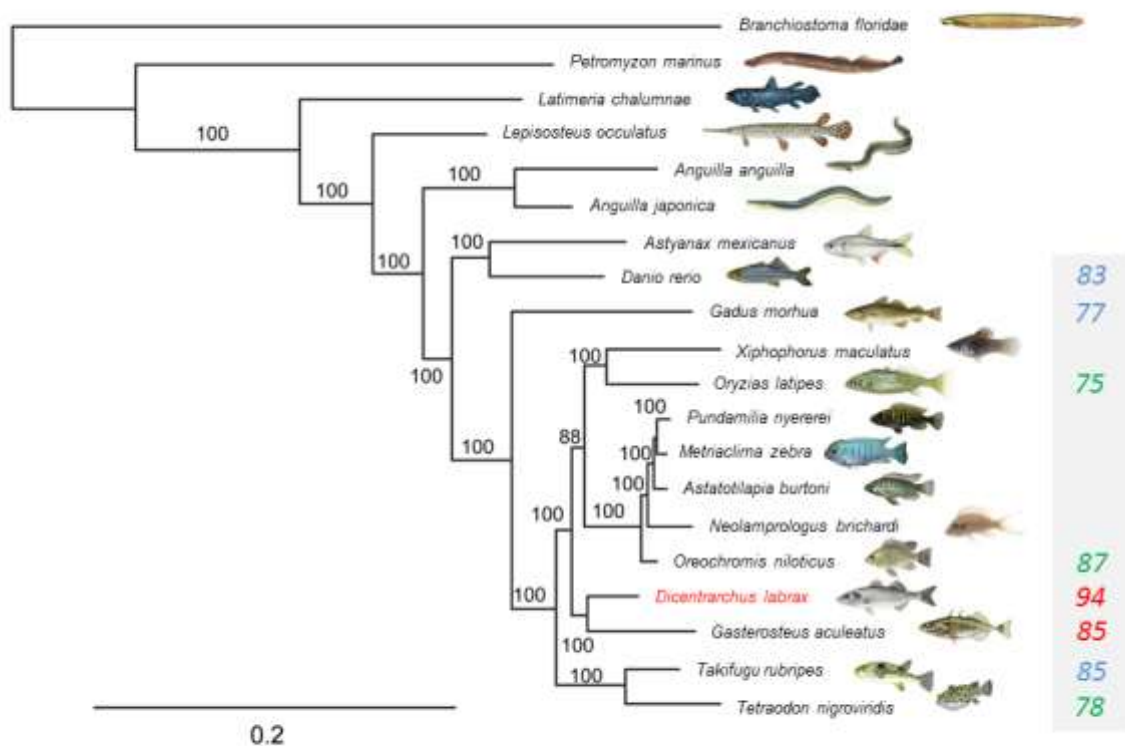


Figure 3: Comparison of claudin gene synteny between sea bass LG 13 and 14 and other vertebrate chromosomes including human. Chromosome blocks containing claudin genes (2) are blue coloured, other synteny blocks are coloured in red (1), yellow (3), orange (4M) and green (4F).

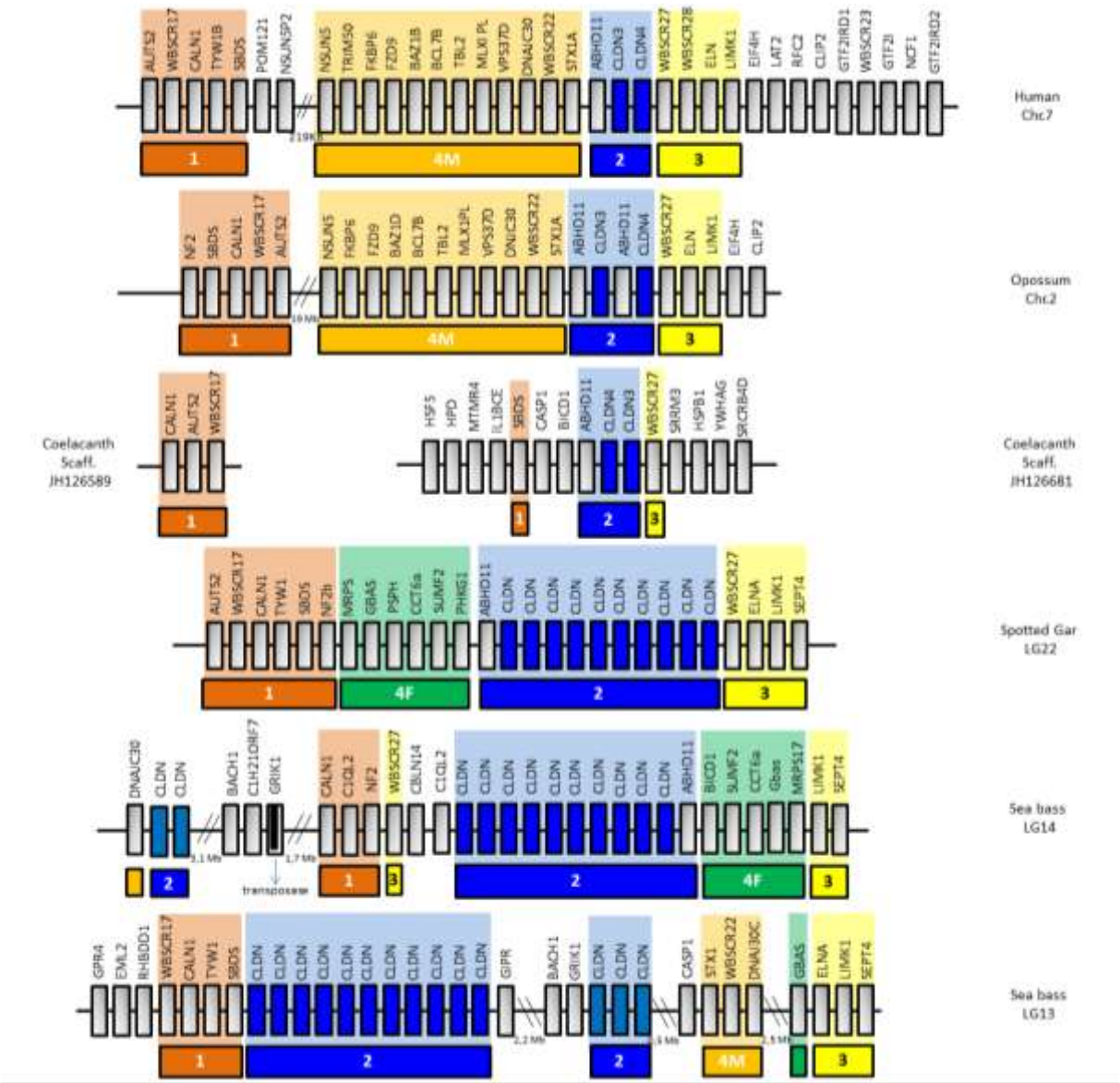
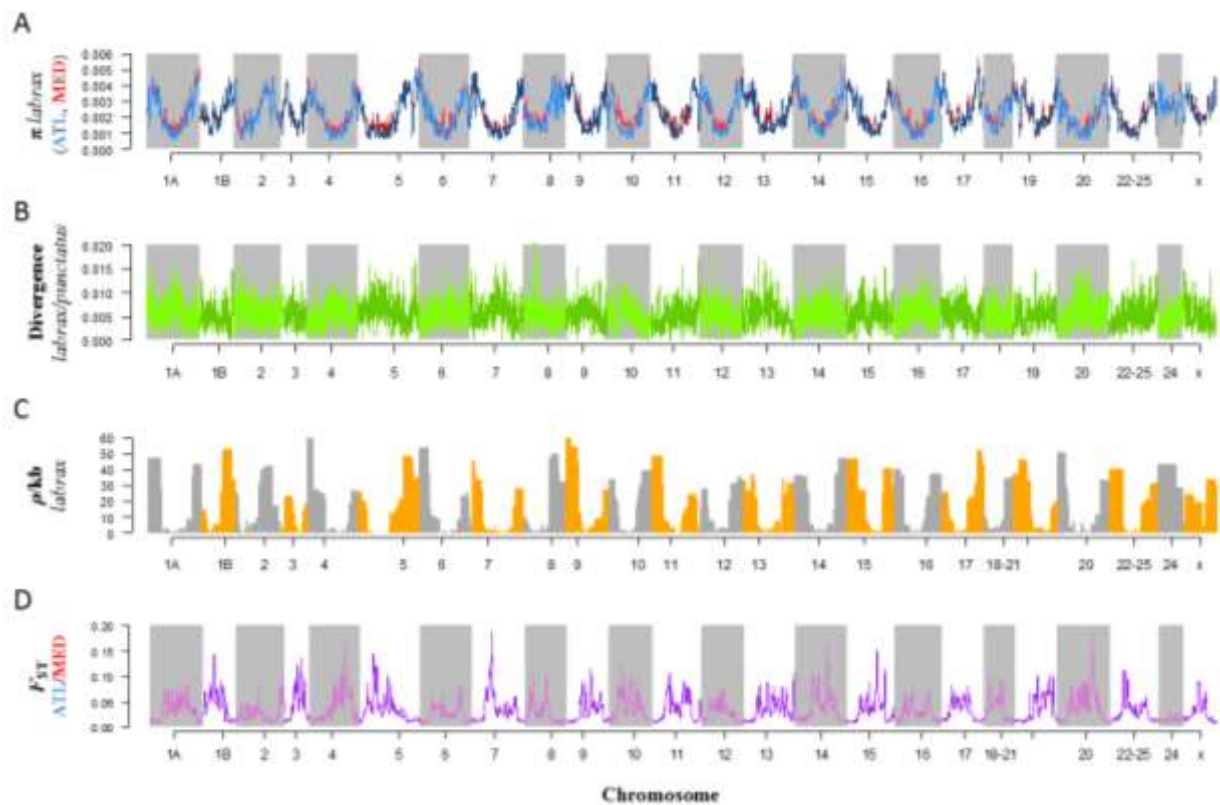


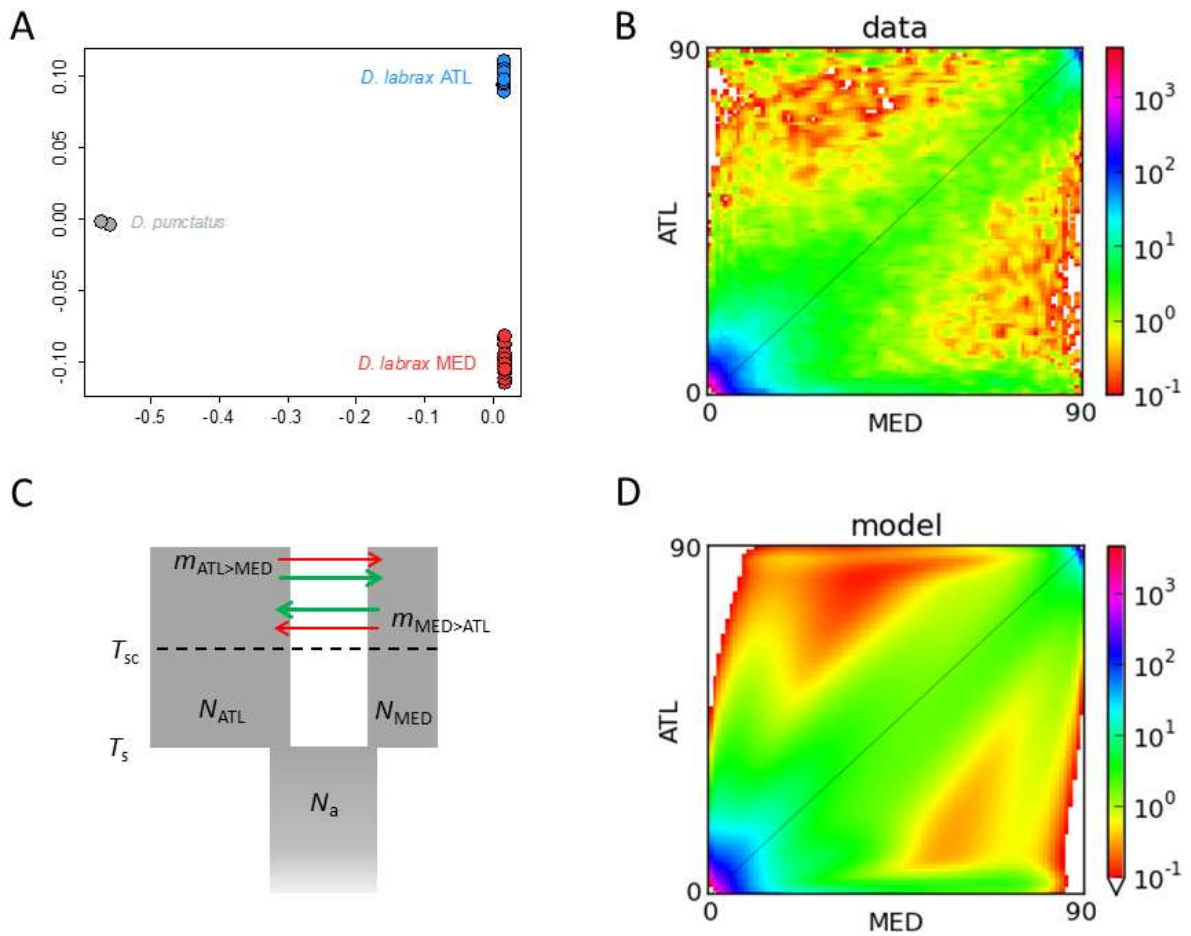
Figure 4: Distribution of population genetic parameters calculated in 150 kb windows across the different chromosomes of sea bass genome (x refers to LG_X, and not to a sexual chromosome). **A:** Chromosomal patterns of nucleotide diversity (π) in the Atlantic (blue) and Mediterranean (red) lineages of *D. labrax*. **B:** Genomic patterns of divergence between *D. labrax* and *D. punctatus* measured as the proportion of fixed differences per bp (d_f). **C:** Chromosomal patterns of recombination (ρ /kb) in *D. labrax*, averaged across Atlantic and Mediterranean lineages, alternately coloured in grey and orange across chromosomes. **D:** Genomic landscape of differentiation (F_{ST}) between Atlantic and Mediterranean lineages.



786

787 **Figure 5:** Population structure and demographic history of the European sea bass,
 788 *Dicentrarchus labrax*. **A:** Principal component analysis of the Atlantic (ATL) and
 789 Mediterranean (MED) sea bass lineages including the outgroup species *D.*
 790 *punctatus*. **B:** The joint allele-frequency spectrum (AFS) for the ATL and MED
 791 populations, showing the count of derived allele for 109,300 oriented SNPs in 45
 792 individuals from each population. **C:** The secondary contact with two migration rate
 793 model (*SC2m*), including 10 parameters: the ancestral population size (N_a), the ATL
 794 and MED population sizes after splitting (N_{ATL} and N_{MED}), the splitting time (T_s) and
 795 the time since secondary contact (T_{SC}). Locus effective migration rate from ATL into
 796 MED ($m_{ATL>MED}$) and in the opposite direction ($m_{MED>ATL}$) can take two free values in
 797 proportions p and $1-p$. **D:** The maximum-likelihood AFS obtained under the *SC2m*
 798 model.

799



800

